数理解析研究所講究録
第 697 巻 1989 年 40-56

ρ-Discrete Languages

H.J. Shyr, G. Thierrin and S.S. Yu

§1. Introduction and Notations

Let X^* be the free monoid generated by the finite alphabet X with $|X| \geq 2$. Any element of X^* is called a *word* and any subset of X^* is called a *language*. The length of a word u is denoted by $lg(u)$. If 1 is the empty word, then $X^+ = X^* \backslash \{1\}$. The *catenation* of two languages A and B is the set $AB = \{xy \mid x \in A, y \in B\}$. A word $u \in X^+$ is *primitive* if $u = f^n, f \in X^+$ implies $n = 1$. Every word can be expressed uniquely as a power of a primitive word ([3]). The set of all primitive words over X will be denoted by Q. If $u = f^n, f \in Q$, then $\sqrt{u} = f$ and for any language $L \subseteq X^+$, $\sqrt{L} = \{\sqrt{u} \mid u \in L\}$.

A nonempty language $L \subseteq X^+$ is called a *code* if $x_1x_2...x_n = y_1y_2...y_m, x_i, y_j \in L$ implies $m = n$ and $x_i = y_i, i = 1, 2, ..., n$ and an *n-code* if every subset of L with at most n elements is a code ([1]).

A language $L \subseteq X^*$ is said to be *n-discrete*, n a positive integer, if $|L \cap X^m| \leq n$ for all $m \geq 1$. L is called *semidiscrete* if L is $n-$discrete for some $n \geq 1$. If $n = 1$, then the language L is said to be *discrete*.

Remark that a language L is n-discrete iff $|L \cap A| \leq n$ for every class A of the equivalence relation λ defined by $u \equiv v(\lambda)$ iff $lg(u) = lg(v)$, because the classes of λ are the sets $\{X^m \mid m \geq 0\}$. It is therefore natural to consider generalizations of the discrete languages in relation with more general equivalence relations ρ.

The purpose of this paper is to study in particular generalizations connected with equivalences associated with general and cyclic permutations of words in X^*. If ρ is an equivalence relation defined on X^*, then the equivalence class containing the word u will be denoted by ρ_u. If n is a positive integer, then $L \subseteq X^*$ is said to be $\rho(n)$-*discrete* if

$$|L \cap \rho_u| \leq n$$

for every $u \in X^*$.

If $n = 1$, then L is called a ρ-*discrete* language.

If $u \in X^*$, then $\pi(u)$ and $\sigma(u)$ denotes respectively the set of all permutations and the set of all cyclic permutations of the word u. The following relations defined on X^* are equivalence relations:

$$(1) \qquad u \equiv v(\lambda) \qquad iff \qquad lg(u) = lg(v);$$

$$(2) \qquad u \equiv v(\sigma) \qquad iff \qquad \sigma(u) = \sigma(v).$$

It is immediate that

$$\sigma \subseteq \pi \subseteq \lambda.$$

It follows then that a $\lambda(n)$-discrete language is a $\pi(n)$-discrete language and that a $\pi(n)$-discrete language is a $\sigma(n)$-discrete language. The converse is not true. For example, if $X = \{a, b\}$, then $\{a^2, ab\}$ is π-discrete, but not λ-discrete and $\{abab, a^2b^2\}$ is σ-discrete but not π-discrete.

Remark that the σ-equivalence classes are the cyclic permutations of a word $u \in X^*$. Hence $\sigma(n)$-discrete languages are the languages containing at most n words of the cyclic permutations $\sigma(u)$ of $u \in X^*$ and a $\sigma(n)$-discrete language is a union of at most n σ-discrete languages.

It is immediate that a language is σ-discrete iff $xy \in L$ and $yx \in L$ implies $xy = yx$. Since a language is called *reflective* iff $xy \in L$ implies $yx \in L$, it follows that a σ-discrete language is, in some way, the opposite of a reflective language and for this reason could also be called an *anti-reflective language.*

In this paper we give, in section 2, several characterizations of $\sigma(n)$-discrete and $\pi(n)$-discrete languages. In section 3, some operations on these two classes of languages are considered and in section 4, the corresponding maximal languages are studied. The special family of σ-discrete 2-codes is investigated in section 5. In the last section, we consider the class of cm-free languages which are in some way the opposite of commutative languages.

§2. Some Properties of $\sigma(n)$-discrete and $\pi(n)$-discrete languages

For any language $L \subseteq X^*$, we let $L^{(m)} = \{x^m \mid x \in L\}$. Clearly if L is a $\sigma(n)$-discrete language, then so is $L^{(m)}$ for $m \geq 2$.

First we establish some characteristic properties of $\sigma(n)$-discrete and $\pi(n)$-discrete languages.

PROPOSITION 2.1. *Let X be an alphabet with $|X| \geq 2$ and let $L \subseteq X^*$. Then for every $n \geq 1$, the following properties are equivalent:*

(1) *L is a $\sigma(n)$-discrete language;*

(2) *$|\sigma(w) \cap L| \leq n$ for all $w \in X^+$;*

(3) *$L \cap X^m$ is $\sigma(n)$-discrete $\forall m \geq 1$;*

(4) *$L^{(m)}$ is $\sigma(n)$-discrete $\forall m \geq 1$;*

(5) *$L^{(m)}$ is $\sigma(n)$-discrete for some $m \geq 1$.*

PROOF. The equivalences of (1),(2) and (3) are immediate.

(1) $\Rightarrow$ (4). Let $m \geq 2$. Suppose $L^{(m)}$ is not $\sigma(n)$-discrete. Then there exist at least $n+1$ distinct words $u_1, u_2, ..., u_n, u_{n+1} \in L$ such that

$$u_i^m \in \sigma(u_1^m) \text{ for all } i.$$

Let $u_i^m = x_i u_1^{m-1} y_i$. Then $x_i \neq 1, y_i \neq 1$ and $u_1 = y_i x_i$. This means that $u_i \in \sigma(u_1)$ for $1 < i \leq n+1$. Thus L is not $\sigma(n)$-discrete, a contradiction.

(4)$\Rightarrow$ (5). Trivial.

(5) $\Rightarrow$ (1). Suppose L is not $\sigma(n)$-discrete. Then there exist at least $n+1$ distinct words $u_1, u_2, ..., u_n, u_{n+1} \in L$ such that $u_i \in \sigma(u_1)$ for all i. It then follows that

$u_i^m \in L^{(m)}$ for all i.

But $u_i \in \sigma(u_1)$ implies that $u_i^m \in \sigma(u_1^m)$. Thus $L^{(m)}$ is not $\sigma(n)$-discrete. This shows that (5) implies (1).◇

A language $L \subseteq X^+$ is called an *infix code* if for $u \in X^+, x, y \in X^*$, $u, xuy \in L$ implies $xy = 1$.

For the case $n = 1$, we have the following proposition:

PROPOSITION 2.2. *Let X be an alphabet such that $|X| \geq 2$ and let $L \subseteq X^*$. Then L is σ-discrete if and only if for any $u, v \in L \cap X^m, \{u^2, v\}$ is an infix code.*

PROOF. Let $u = a_1 a_2 ... a_m$; $v = b_1 b_2 ... b_m$, where $a_i, b_j \in X$, Then $\{u^2, v\}$ is not an infix code if and only if $u^2 = xvy$ for some $x, y \in X^*, xy \neq 1$. Which then implies that $\{u^2, v\}$ is not an infix code if and only if $v = a_i a_{i+1} ... a_m a_1 a_2 ... a_{i-1}$, for some $1 \leq i \leq m$. The proof of the proposition follows then easily from these results.◇

PROPOSITION 2.3. *Let X be an alphabet such that $|X| \geq 2$. Let $L \subseteq X^*$. Then for any $n \geq 1$, the following properties are equivalent*

(1) *L is a $\pi(n)$-discrete language;*

(2) *$|\pi(w) \cap L| \leq n$ for all $w \in X^+$;*

(3) *$L \cap X^m$ is $\pi(n)$-discrete $\forall m \geq 1$;*

(4) *$L^{(m)}$ is $\pi(n)$-discrete $\forall m \geq 1$;*

(5) *$L^{(m)}$ is $\pi(n)$-discrete for some $m \geq 1$.*

PROOF. The equivalences of (1),(2) and (3) are immediate.

(1) $\Rightarrow$ (4). Let $m \geq 2$. Suppose $L^{(m)}$ is not $\pi(n)$-discrete. Then there exist at least $n+1$ distinct words $u_1, u_2, ..., u_n, u_{n+1} \in L$ such that

$$u_i^m \in \pi(u_1^m) \text{ for all } i.$$

This means that $u_i \in \pi(u_1)$ for $1 < i \leq n+1$. Thus L is not $\pi(n)$-discrete, a contradiction.

(4)$\Rightarrow$ (5). Trivial.

(5) $\Rightarrow$ (1). Suppose L is not $\pi(n)$-discrete. Then there exist at least $n+1$ distinct words $u_1, u_2, ..., u_n, u_{n+1} \in L$ such that $u_i \in \pi(u_1)$ for all i. It then follows that $u_i^m \in L^{(m)}$ for all i. But $u_i \in \pi(u_1)$ implies that $u_i^m \in \pi(u_1^m)$. Thus $L^{(m)}$ is not $\pi(n)$-discrete. This shows that (5) implies (1).$\Diamond$

It has been shown that a semi-discrete dense language is disjunctive (see [2]). The following proposition is a generalization of this fact.

PROPOSITION 2.4. *A $\pi(n)$-discrete language L is dense if and only if L is disjunctive.*

PROOF. ($\Leftarrow$) Trivial.

($\Rightarrow$) Let L be a $\pi(n)$-discrete dense language. For any $u \neq v \in X^*$, there exist $x, y \in X^*$ such that $\sqrt{xuy} \neq \sqrt{xvy}$. Let $u' = xuyxvy$ and let $v' = xvyxuy$. Then $u' \in \pi(v')$. Define $w_1 = (u')^n$, $w_2 = (u')^{n-1}v'$,...,$w_n = u'(v')^{n-1}$, $w_{n+1} = (v')^n$. Then $w_i \in \pi(w_1)$ for all i. If $u \equiv v(P_L)$, then $u' \equiv v'(P_L)$. This implies that $w_i \equiv w_j(P_L)$ for all i, j. Since L is dense, there exist $z, z' \in X^*$ such that $zw_1z' \in L$. This implies that $zw_iz' \in L$ for all i. But $zw_iz' \in \pi(zw_1z')$ for all i and this contradicts the condition that L is $\pi(n)$-discrete. Thus $u \not\equiv v(P_L)$ for all $u \neq v \in X^*$. This shows that L is disjunctive.$\Diamond$

§3. Operations on σ-discrete and π-discrete Languages

For a language $L \subseteq X^*$, let $L^c = X^* \backslash L$ be the complement of L in X^*.

PROPOSITION 3.1. *Let ρ be an equivalence relation such that $\sigma \subseteq \rho$. Then for any $\rho(n)$-discret language L, L^c is dense.*

PROOF. Since every $\rho(n)$-discrete language is a $\sigma(n)$-discrete language, we only need to show that for any $\sigma(n)$-discrete language L, L^c is dense.

Now let L be a $\sigma(n)$-discrete language and suppose L^c is not dense. Then there exists a word $w \in X^+$ such that $X^*wX^* \cap L^c = \emptyset$. It then implies that $X^*wX^* \subseteq L$. Let $m = lg(w)$. Then $|\sigma(w^2ab^{2m+n}a)| > n$. This contradicts the condition that L is $\sigma(n)$-discrete. Therefore, L^c must be dense.

COROLLARY 3.2. *For any $\pi(n)$-discrete language L, L^c is dense.*$\diamondsuit$

It is clear that if L is not a σ-discrete language, then L^i is not σ-discrete for all $i \geq 2$. If L is σ-discrete, then L^i is not necessarily σ-discrete. In fact, the next proposition shows that, for example, the class of languges L such that L and L^2 are σ-discrete is quite restrictive.

PROPOSITION 3.3. *Let L be a language. Then the following properties are equivalent:*

(1) *L and L^2 are σ-discrete;*

(2) *L and L^2 are π-discrete;*

(3) *$L \subseteq w^*$ for some $w \in X^*$.*

PROOF. (2) $\Rightarrow$ (1) Since every π-discrete language is σ-discrete, clearly (2) implies (1).

(1) $\Rightarrow$ (3) Suppose $L \not\subseteq w^*$ for any $w \in X^*$. Then there exist $x, y \in L$ such that $x \neq 1 \neq y$ and $\sqrt{x} \neq \sqrt{y}$. Since $xy \neq yx$ and $xy, yx \in L^2$, L^2 is not σ-discrete, a contradiction.

(3) $\Rightarrow$ (2) Suppose $L \subseteq w^*$ for some $w \in X^*$. Then clearly $L^2 \subseteq w^*$ and L^2 is discrete. Therefore, L and L^2 are π-discrete.$\diamondsuit$

In relation with the preceding proposition, we have the following result:

PROPOSITION 3.4. *Let $L \subseteq X^*$. Then the following properties are equivalent:*

(1) *L is a σ-discrete submonoid;*

(2) *L is a π-discrete submonoid;*

(3) *$L = w^*$ for some $w \in X^*$.*

PROOF. (2) $\Rightarrow$ (1) Since every π-discrete language is a σ-discrete language, the implication holds.

(1) $\Rightarrow$ (3) Suppose there exist $w_1, w_2 \in L$ with $w_1 \neq 1$ and $w_2 \neq 1$ such that $\sqrt{w_1} \neq \sqrt{w_2}$. Then $w_1w_2 \neq w_2w_1$ and $w_1w_2, w_2w_1 \in L$. Which implies that L is not σ-discrete, a contradiction. Therefore, $L = w^*$ for some $w \in X^*$.

(3) $\Rightarrow$ (2) Trivial.◇

In general, if a language L is σ-discrete then $\sqrt{L}$ is not necessarily σ-discrete. For example, $L = \{a^2b, (aba)^2\}$ is σ-discrete but $\sqrt{L} = \{a^2b, aba\}$ is not. However the converse is true for any language $L \subseteq X^+$.

PROPOSITION 3.5. *Let $L \subseteq X^+$. If $\sqrt{L}$ is a σ-discrete language, then L is σ-discrete.*

PROOF. Suppose L is not σ-discrete. Then there exist $u, v \in L$ such that $u \in \sigma(v)$ and $u \neq v$. Let $v \in Q^{(i)}$ for some i. Then by Proposition 1.11 ([7]), $u \in Q^{(i)}$. Thus $v = g^i$ and $u = h^i$ for some $g \neq h \in Q$. Which then implies that $h \in \sigma(g)$ and $g, h \in \sqrt{L}$. Thus $\sqrt{L}$ is not σ-discrete, a contradiction. Therefore, L is σ-discrete.◇

The next proposition shows that the family of σ-discrete languages is not closed under catenation.

PROPOSITION 3.6. *For any word $w \in X^+$, there exists a σ-discrete language L such that wL is not σ-discrete .*

PROOF. Let $X = \{a, b, ...\}$. Given $w \in X^+$:

(i) if $w \notin b^+$, then we let $L = \{bw^3b, bwbw^2\}$;

(ii) if $w = b^n, n \geq 1$, then we let $L = \{aw^3a, awaw^2\}$. It is clear that L is σ-discrete but wL is not. This proves the proposition.$\diamondsuit$

In the following discussion, we consider the free monoid X^* with the *standard total order* $\leq$ which is defined as follows (see [6]):
For $u, v \in X^*$, $u < v$ if $lg(u) < lg(v)$ and $\leq$ is the lexicographical order if $lg(u) = lg(v)$.
Let $A = \{a_1 < a_2 < ... < a_i < ...\}$ and $B = \{b_1 < b_2 < ... < b_i < ...\}$ be two languages over X with the same cardinality and ordered relatively to the standard order. The *ordered catenation of A and B* is the set

$A\Delta B = \{a_i b_i \mid i = 1, 2, ...\}$.

We let $A^{(2)} = A\Delta A$ and let $A^{(n)} = A^{(n-1)}\Delta A$ for $n \geq 3$. Let $(X^*, \leq) = \{x_0 < x_1 < x_2 < ... < x_i < ...\}$ with the standard total order $\leq$. The injective mapping $\# : X^* \mapsto \mathrm{N}\cup\{0\}$ is defined by $\#(x) = i$ if $x = x_i$.

In general, σ-discrete languages are not closed under ordered catenation. In the next proposition, we consider a case where this is true. For the proof of this proposition, we need the following known results:

(*) If u and v have powers u^m and v^n with a common initial segment of length $lg(u) + lg(v)$, then u and v are powers of a common word ([3]).

In particular we have

(**) For $p, q \in Q$, if p^i and q^j have a common segment of length $lg(p) + lg(q)$, then $p \in \sigma(q)$.

PROPOSITION 3.7. *Let $A \subseteq Q^{(i)}, B \subseteq Q^{(j)}$ where $i \neq j \geq 3$. If both $(A, \leq)$ and $(B, \leq)$ are σ-discrete, then $A\Delta B$ is σ-discrete.*

PROOF. Suppose $A\Delta B$ is not σ-discrete. Then there exist $u_1, u_2, v_1, v_2 \in Q$ with $u_1^i, u_2^i \in A$, $v_1^j, v_2^j \in B$, $u_1^i v_1^j, u_2^i v_2^j \in A\Delta B$ and $u_1^i v_1^j \in \sigma(u_2^i v_2^j)$. Which implies that $lg(u_1^i v_1^j) = lg(u_2^i v_2^j)$, $lg(u_1) = lg(u_2), lg(v_1) = lg(v_2)$. Since both A and B are σ-discrete, $u_1 \notin \sigma(u_2)$, $v_1 \notin \sigma(v_2)$. Thus $u_1^i \notin E(\sigma(u_2^i))$, $v_1^j \notin E(\sigma(v_2^j))$ and vice versa. It is clear that

$lg(u_k^i) < lg(u_1) + lg(v_1^j)$ and $lg(v_k^j) < lg(v_1) + lg(u_1^i)$ for $k = 1, 2$. (Otherwise, $u_1 \in \sigma(u_2)$ or $v_1 \in \sigma(v_2)$.) Without loss of generality, let $\#(u_1) > \#(u_2)$ and let $\#(v_1) > \#(v_2)$. Then we have the following five cases.

Case 1, $u_2^i = xy, v_1^j = yz$ for some $x, y, z \in X^*$ with $lg(x) < lg(u_2)$ and $lg(z) < lg(v_1)$. Then $lg(y) > 2\max\{lg(u_2), lg(v_1)\} \geq lg(u_2) + lg(v_1)$. By the condition (**) above, $u_2 \in \sigma(v_1)$. Thus $lg(u_2) = lg(v_1)$. This implies that $i = j$; a contradiction.

Case 2, $v_1^j = xy, u_2^i = yz$. It is the same as Case 1.

Case 3, $u_2^i = xv_2^jy$ and $lg(x) + lg(y) < lg(u_2)$. Since $i, j \geq 3$, $lg(v_1^j) > 2\max\{lg(u_2), lg(v_1)\}$. By the condition (**) above, $u_2 \in \sigma(v_1)$. We get that $lg(u_2) = lg(v_1)$ and $i = j$; a contradiction.

Case 4, $v_1^j = xu_2^iy$. It is the same as Case 3.

Case 5, $v_1^j = u_2^i$. By the condition (*) above, $u_2 = v_1$. Then $u_1 = v_2$. But $\#(v_2) = \#(u_1) > \#(u_2) = \#(v_1)$; a contradiction.

Therefore, the language $A \triangle B$ must be σ-discrete.$\diamondsuit$

Let A and B be two σ-discrete languages contained in $Q^{(i)}, Q^{(j)}$ respectively. If $i = j$, then $A \triangle B$ may not be σ-discrete. This is the case, for example, if $A = \{(aaba)^i, (bbaa)^i\}$ and $B = \{(aabb)^i, (baaa)^i\}$. Then $A, B \subseteq Q^{(i)}$ and both A and B are σ-discrete. However, $A \triangle B$ is not σ-discrete. If $A \subseteq Q$, then $A \triangle B$ may also not be σ-discrete. For example, let $A = \{ab^jbb, ba^jab\} \subseteq Q$ and let $B = \{a^j, b^j\} \subseteq Q^{(j)}$. Then both A and B are σ-discrete. But $A \triangle B$ is not σ-discrete.

§4. Maximal σ-discrete and π-discrete Languages

DEFINITION. An σ-discrete language $L \subseteq X^+$ is *maximal* if L is not properly contained in other σ-discrete languages, that is, for any σ-discrete language $L' \subseteq X^+$, $L \subseteq L'$ imples that $L = L'$.

PROPOSITION 4.1. *Let $L \subseteq X^+$. Then the following properties are equivalent:*

(1) *L is a maximal σ-discrete language;*

(2) $|L \cap \sigma(w)| = 1$ *for all* $w \in X^+$*;*

(3) $L \cap X^i$ *is a maximal σ-discrete language in* X^i, $i \geq 1$*;*

(4) $L \cap Q^{(i)}$ *is a maximal σ-discrete language in* $Q^{(i)}$, $i \geq 1$.

PROOF. Immediate.◇

The elements of a maximal σ-discrete language have the following interesting properties:

If L is a maximal σ-discrete language, then for any $v \in X^+$, there exist some $x, y \in X^*$ such that $yv^ix \in L$ for some i, and there also exist some $x, y \in X^*$ such that $(yvx)^i \in L$ for some i. In fact:

LEMMA 4.2. *Let L be a maximal σ-discrete language. Then for any $v \in X^+$ and for any $i \geq 1$ there exist $x, y \in X^*$ with $xy = v$ such that $(yx)^{i+1} = yv^ix \in L$.*

PROOF. Let $v \in X^+$. Then by Proposition 4.1, $\sigma(v^{i+1}) \cap L \neq \emptyset$. Let $v = xy$ for some $x, y \in X^*$ be such that $v^{i+1} = xyxy...xy$ *and* $yxyx....yx \in L$. Then clearly $yxyx...yx = (yx)^{i+1} = yv^ix \in L$.◇

An immediate result of Lemma 4.2, we have the following:

REMARK 4.3. *If L is a maximal σ-discrete language, then for any $v \in X^+$ and $i \geq 1$ there exist $x, y \in X^*$, $xy = v$ such that $(yvx)^i \in L$.*

Recall that a language L is called *dense* if for any $v \in X^+$, there exist $x, y \in X^*$ such that $xvy \in L$. The language L is called *disjunctive* if its *syntatic congruence* P_L is the equality, where P_L is defined by $u \equiv v(P_L)$ if and only if $L..u = L..v$ with $L..u$ being the set of all pairs of words (x, y) such that $xuy \in L$. Every disjunctive language is dense, but the converse is not true.

By Lemma 4.2 or by the above Remark, a maximal σ-discrete language L is always dense and we will show in the next proposition that it is also disjunctive. However if L is not maximal, then L is not necessarily disjunctive. For example, let $X = \{a, b\}$ and let $L = \{bxba^{lg(x)+2} | x \in X^+\}$. It is clear that L is a σ-discrete and dense language that is not disjunctive.

PROPOSITION 4.4. *Every maximal σ-discrete language is a disjunctive language.*

PROOF. Suppose L is a maximal σ-discrete language which is not disjunctive. Then there exist two words $u, v \in X^+, u \neq v, lg(u) = lg(v)$ such that $u \equiv v(P_L)$. It follows that $(xvy)^2 \equiv xvyxuy \equiv xuyxvy(P_L)$ for all $x, y \in X^*$. By Lemma 4.2 there exist x, y such that $(xvy)^2 \in L$. Which then implies that $xvyxuy \in L$ and $xuyxvy \in L$, a contradiction.$\Diamond$

Let S be any finite set. If γ is a *permutation* of S let $\psi(\gamma) = |\{s \in S | \gamma(s) = s\}|$.

Now, let $S = X^n$ and let γ be the permutation defined by $\gamma(a_1a_2...a_n) = a_2...a_na_1$ where $a_1a_2...a_n \in S$. Then clearly $\gamma^n(x) = x$ for all $x \in S$. Thus, γ^n stands as unit element of G where $G = \{\gamma, \gamma^2, \gamma^3, ..., \gamma^n\}$, and $\psi(\gamma^n) = |X^n|$. Two elements s_1, s_2 of S are called *equivalent*, written $s_1 \sim s_2$, if there exists a permutation $\gamma^i \in G$ such that $\gamma^i(s_1) = s_2$. It is clear that $\sim$ is an equivalence relation. For $\gamma^i \in G$, the order of γ^i is the least positive integer k such that $(\gamma^i)^k = \gamma^n$. Hence, the order of γ^n is 1.

Let ϕ be the Euler's function; that is, $\phi(d)$ is the number of positive integers k with $1 \leq k \leq d$, $(k, d) = 1$. Then, by [4], we have the following result:

$$|S/\sim| = \frac{1}{|G|}\Sigma_{\gamma\in G}\psi(\gamma)$$
$$= \frac{1}{n}\Sigma_{d|n}\psi(\gamma_d)\phi(\tfrac{n}{d})$$

where $\gamma_d \in G$ and the order of γ_d is d.

Hence for any maximal σ-discrete language L, we can calculate the number of elements in the intersection of L and X^n with the following formula:

$$(\alpha)\ |L \cap X^n| = |X^n/\sim| = \frac{1}{n}\Sigma_{\gamma\in G}\psi(\gamma).$$

If $L \subseteq X^*$ and if $|L \cap X^n| \leq cn$ for some constant c, then L is called *linear discrete*. Using the formula (α) showed above, we now prove that every maximal σ-discrete language over a finite alphabet X is not linear discrete.

PROPOSITION 4.5. *Let $|X| = k \geq 2$. Then every maximal σ-discrete language over X is not linear discrete.*

PROOF. Since $|X| = k$, then $|X^n| = k^n$. Let L be a maximal σ-discrete language over X. By formula (α), $|L \cap X^n| = \frac{1}{n}\Sigma_{\gamma \in G}\psi(\gamma)$. But $\lfloor \frac{k^n}{n} \rfloor \leq \frac{1}{n}\Sigma_{\gamma \in G}\psi(\gamma)$ and $lim_{n\to\infty}\lfloor \frac{k^n}{n^2} \rfloor \to \infty$. Thus there exists no constant c such that $|L \cap X^n| \leq cn$. Therefore L is not linear discrete.$\diamondsuit$

Let $X = \{a_1, a_2, ..., a_k\}$. Then the language $L = a_1^* a_2^* ... a_k^*$ is a maximal and regular π-discrete language. It is clear that every maximal π-discrete language has the same number of elements in X^n, we need only to consider $|L \cap X^n|$. From [5], we know that $|L \cap X^n|$ is equal to the conbination number $\mathrm{C}\binom{k+n-1}{n} = \frac{(k+n-1)!}{n!(k-1)!}$. Hence:

REMARK 4.6. *Let $|X| = k$ and let L be a maximal π-discrete language. Then* $|L \cap X^n| = \mathrm{C}\binom{k+n-1}{n}$.

Now we show that a maximal π-discrete language is not linear discrete.

PROPOSITION 4.7. *Let $|X| = k \geq 2$. Then every maximal π-discrete language is not linear discrete.*

PROOF. By the above Remark, we know that $|L \cap X^n| = \frac{(k+n-1)!}{n!(k-1)!}$ for any maximal π-discrete language L. Since $\lim_{n\to\infty}\frac{1}{n}(\frac{(k+n-1)!}{n!(k-1)!}) = \infty$, L is not linear discrete.$\diamondsuit$

§5. σ-discrete 2-Codes

An *σ-discrete 2-code* is a σ-discrete language which is also a 2-code. For any $i \geq 1$, every σ-discrete language contained in $Q^{(i)}$ is such a language.

PROPOSITION 5.1. *Let $L \subseteq X^+$. Then L is an σ-discrete 2-code if and only if for every $v = f^i, f \in Q$ $i \geq 1$,*

(i) $|f^+ \cap L| \leq 1$,

(ii) if $f^r \in L$, then $g^r \notin L$ for all $g \in \sigma(f)$ and $g \neq f$.

PROOF. Immediate.◇

We call a language $L \subseteq X^+$ a *maximal σ-discrete 2-code* if for every σ-discrete 2-code L' such that $L \subseteq L'$, then $L = L'$. In Proposition 4.4 it was proved that every maximal σ-discrete language is disjunctive. The following proposition shows that this is also true for every maximal σ-discrete 2-code.

PROPOSITION 5.2. *If L is a maximal σ-discrete 2-code, then L is disjunctive.*

PROOF. Let $L \subseteq X^+$ be a maximal σ-discrete 2-code. Suppose for some $u \neq v \in X^n$, $n \geq 1$ such that $u \equiv v(P_L)$. Clearly, $u^2v^2 \in Q$.

Suppose $\sigma(u^2v^2) \cap L \neq \emptyset$. We have two cases:

(i) there exist $x, y \in X^*$, $xy = v$ such that $yvu^2x \in L$ or $yu^2vx \in L$;

(ii) there exist $x, y \in X^*$, $xy = u$ such that $yuv^2x \in L$ or $yv^2ux \in L$.

Since $u \equiv v(P_L)$, $u^2v \equiv vu^2(P_L)$ and $uv^2 \equiv v^2u(P_L)$ hold. This in turns implies that $yu^2vx \in L \iff yvu^2x \in L$ and $yuv^2x \in L \iff yv^2ux \in L$. From this fact and since L is σ-discrete, we see that $\sigma(u^2v^2) \cap L = \emptyset$ must be true. Now, (1) if $(u^2v^2)^i \notin L$ for all $i \geq 1$, then $L \cup \{u^2v^2\}$ is an σ-discrete 2-code and which contradict to the maximality of L. (2) If there exists an $i \geq 2$ such that $(u^2v^2)^i \in L$, then since

$$(u^2v^2)^i \equiv uv^2(u^2v^2)^{i-1}u(P_L)$$

$uv^2(u^2v^2)^{i-1}u \in L$ holds, a contradiction. This shows that every maximal σ-discrete 2-code is a disjunctive language.◇

Recall that $Q^{(i)}$ is a maximal 2-code and that every σ-discrete language contained in $Q^{(i)}$ for any $i \geq 1$ is a σ-discrete 2-code. However such a language cannot be a maximal σ-discrete 2-code:

PROPOSITION 5.3. *For any $i \geq 1$, there exists no maximal σ-discrete 2-code contained in $Q^{(i)}$.*

PROOF. Suppose on the contrary that there is a maximal σ-discrete 2-code $L \subseteq Q^{(i)}$ for some $i \geq 1$. Then $(ab)^i \in L$ or $(ba)^i \in L$, $a \neq b \in X$. Indeed, if $(ab)^i \notin L$ and $(ba)^i \notin L$, then $L \cap (ba)^i$ is a σ-discrete language contained in $Q^{(i)}$ and L is not a maximal σ-discrete 2-code contaied in $Q^{(i)}$. Now let us assume $(ab)^i \in L$. Since L is a 2-code, $(ab)^{i+1} \notin L$. Again since $L \subseteq Q^{(i)}$, we have $(ba)^j \notin Q^{(i)}$ for all $j \geq 1$. It then follows that $L \cap \{(ba)^{j+1}\}$ is a σ-discrete 2-code. This implies that L is not a maximal σ-discrete 2-code, a contradiction. This shows that for $i \geq 1$, $Q^{(i)}$ contains no maximal σ-discrete 2-code.$\diamondsuit$

We give now a method to construct maximal σ-discrete 2-codes.

Let $A \in X^+$ be a non empty language. A σ-discrete language $L \subseteq A$ is called *A-maximal* if there is no σ-discrete language in A containing strictly L. Since every non empty word is a σ-discrete language, then, by the Zorn's Lemma, A always contains a A-maximal σ-discrete language. For a language $L \subseteq X^+$, $L^{(+)}$ denotes the set $L^{(+)} = \bigcup_{x \in L}(\sqrt{x})^+$.

We construct a sequence of languages $L_1, L_2, L_3, ...$ inductively in the following way: First we choose a Q-maximal σ-discrete language L_1 in Q. This is always possible by the above considerations and L_1 is a 2-code. Let $T_2 = Q^{(2)} - L_1^{(+)}$.
Next we choose a T_1-maximal σ-discrete language L_2 in T_2. The language $L_1 \cup L_2$ is also a 2-code. Let $T_3 = Q^{(3)} - (L_1^{(+)} \cup L_2^{(+)})$. Suppose now that we have chosen the language L_n which is a T_n-maximal σ-discrete language in

$$T_n = Q^{(n)} - (L_1^{(+)} \cup L_2^{(+)} \cup ... L_{n-1}^{(+)}).$$

We choose then a T_{n+1}-maximal σ-discrete language L_{n+1} in $T_{n+1} = Q^{(n+1)} - (L_1^{(+)} \cup L_2^{(+)} \cup ... \cup L_n^{(+)})$.

By induction, we have now a sequence of languages $L_1, L_2, L_3, ...$ that are disjoint σ-discrete 2-codes. Let

$$L = \bigcup_{n=1}^{\infty} L_n.$$

It is easy to see that the language L is a maximal σ-discrete language which is also a maximal 2-code. It follow then that L is a maximal σ-discrete 2-code.

§6. cm-free languages

A language $L \subseteq X^*$ is said to be *commutative* or *abelian* if for all $u, v, x, y \in X^*$, $yuvx \in L \iff yvux \in L$. This is equivalent to the property that the syntactic monoid of L is a commutative monoid. For the properties of abelian regular languages, see for example ([7]). A language L is called $cm\text{-}free$ or $commutativity - free$ if $xuvy \in L$ and $u \neq v$, $x, u, v, y \in X^*$, implies $xvuy \notin L$. For example, the language $L = a^+ \cup b^+$ with $a \neq b \in X$ is a cm-free language. It is immediate that a *cm-free language is σ-discrete.* It is also clear that every *discrete language is cm-free.* For dense cm-free languages, we have the following:

PROPOSITION 6.1. *Every cm-free language $L \subseteq X^*$ that is dense, is disjunctive.*

PROOF. Suppose that L is dense but not disjunctive. Then there exist $u \neq v \in X^*$ such that $u \equiv v(P_L)$. It is possible to find a word w such that both uw and vw are primitive. Since P_L is a congruence, then $uw \equiv vw(P_L)$ and $uwvw \equiv vwuw(P_L)$ with $uwvw \neq vwuw$. Since L is dense, there exist $x, y \in X^*$ such that $xuwvwy, xvwuwy \in L$. Hence L is not cm-free, a contradiction.$\diamondsuit$

PROPOSITION 6.2. *Every maximal cm-free language is dense and hence disjunctive.*

PROOF. Let L be a maximal cm-free language and let $w \in X^*$. If $w = 1$, then $L \cap X^*wX^* = L \neq \emptyset$. If $w \neq 1$, then we consider the word w^5. Since L is maximal cm-free, there is a word $xuvy \in L$ with $xvuy = w^5$ for some $x, u, v, y \in X^*$. Then w must be a subword of x, u, v or y. This means that $xuvy \in X^*wX^* \cap L \neq \emptyset$. The disjunctivity of L follows from Proposition 6.1.$\diamondsuit$

PROPOSITION 6.3. *For any $x, y \in X^*$, the language $\{x, y\}$ is cm-free if and only if $\{uxv, uyv\}$ is cm-free for all $u, v \in X^*$.*

PROOF. Since $x = w_1w_2w_3w_4, y = w_1w_3w_2w_4$ for some $w_1, w_2, w_3, w_4 \in X^*$ if and only if $uxv = uw_1w_2w_3w_4v, uyv = uw_1w_3w_2w_4v$, $\{x, y\}$ is commutative if and only $\{uxv, uyv\}$ is commutative.$\diamondsuit$

LEMMA 6.4. *For any $x \neq y \in X^*$ there exists a word $w \in X^*$ such that $xwy \neq ywx$.*

PROOF. If $xy \neq yx$, then let $w = 1$. If $xy = yx$, then $lg(x) \neq lg(y)$. Without loss of generality, we can take $lg(x) > lg(y)$. Suppose that $a \in \{u \mid u \in X \textit{ and } yuz = x \textit{ for some } y, z \in X^*\}$. Since $|X| \geq 2$, there exists a $b \in X$ with $b \neq a$. Let $n = lg(x)+lg(y)$ and let $w = b^n$. If $xwy = ywx$, then $x = b^iy = yb^i$ for some i and then $x, y \in b^*$ ([3]). But $a \in \{u \mid u \in X \textit{ and } yuz = x \textit{ for some } y, z \in X^*\}$ and $b \neq a$, a contradiction. Thus $xwy \neq ywx$.$\diamondsuit$

The next proposition shows that in general cm-free languages are not closed under catenation.

PROPOSITION 6.5. *For any cm-free language L with $|L| \geq 2$, there exists a cm-free language L' such that LL' is not cm-free.*

PROOF. Suppose $x, y \in L$ with $x \neq y$. Then by the above Lemma, there exists a word $w \in X^*$ such that $xwy \neq ywx$ and hence $xwyw \neq ywxw$. By Proposition 6.3, $\{wxw, wyw\}$ is a cm-free language. Let $L' = \{wxw, wyw\}$. Then $\{xwyw, ywxw\} \subseteq LL'$ and LL' is not cm-free.$\diamondsuit$

References

[1] Ito, M., Jürgensen, H., Shyr, H.J. and Thierrin, G., *Anti-Commutative Languages and n-Codes*, to appear in Applied Discrete Mathematics.

[2] Kunze, M., Shyr, H.J. and Thierrin, G., *H-bounded and Semi-discrete Languages*, Information and Control, **51** (1981), 174-187.

[3] Lyndon, R.C. and Schützenberger, M.P., *The Equation* $a^M = b^N c^P$ *in a Free Group*, Michgan Math, J. **9** (1962), 289-298.

[4] Pólya, G., *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemishe Verbimdungen,* Acta Math. **68** (1937), 145-254.

[5] Riordan, J., *An Introduction to Combinatorial Analysis*, John Wiley & Sons, Inc., New York, 1958.

[6] Shyr, H.J., *Ordered Catenation and Regular Free Disjunctive Languages,* Information and Control **46** (1980), 257-269.

[7] Shyr, H.J., *Free Monoids and Languages*, Lecture Notes, Department of Mathematics, Soochow University, Taipei, 1979.

Department of Mathematics
The University of Western Ontario
London, Ontario, Canada
N6A 5B7

This research has been supported by the Natural Sciences and Engineering Research Council of Canada under Grant A7877.